المنظومة
ALMANDUMAH

العنوان:	التنبؤ بالمبيعات بإستخدام شجرة القرار: بالتطبيق على شركة سويتش للتجارة والهندسة قسم المولدات
المؤلف الرئيسي:	عبدالله، سمية عبدالله محمد الأمين
مؤلفين آخرين:	أحمد، شاذلى صديق محمد(مشرف)
التاريخ الميلادي:	2017
موقع:	الخرطوم
الصفحات:	1 - 82
رقم MD:	916054
نوع المحتوى:	رسائل جامعية
اللغة:	Arabic
الدرجة العلمية:	رسالة ماجستير
الجامعة:	جامعة النيلين
الكلية:	كلية علوم الحاسوب وتقانة المعلومات
الدولة:	السودان
قواعد المعلومات:	Dissertations
مواضيع:	التنبؤ بالمبيعات، شركة سويتش للتجارة والهندسة، قسم المولدات، تقنية أشجار القرار، التنقيب عن البيانات، قواعد البيانات
رابط:	http://search.mandumah.com/Record/916054

© 2024 المنظومة. جميع الحقوق محفوظة.
هذه المادة متاحة بناء على الإتفاق الموقع مع أصحاب حقوق النشر، علما أن جميع حقوق النشر محفوظة. يمكنك تحميل أو طباعة هذه المادة للاستخدام الشخصي فقط، ويمنع النسخ أو التحويل أو النشر عبر أي وسيلة (مثل مواقع الانترنت أو البريد الالكتروني) دون تصريح خطي من أصحاب حقوق النشر أو المنظومة.

للإستشهاد بهذا البحث قم بنسخ البيانات التالية حسب إسلوب الإستشهاد المطلوب:

إسلوب APA
عبدالله، سمية عبدالله محمد الأمين، و أحمد، شاذلى صديق محمد. (2017).التنبؤ بالمبيعات بإستخدام شجرة القرار: بالتطبيق على شركة سويتش للتجارة والهندسة قسم المولدات(رسالة ماجستير غير منشورة). جامعة النيلين، الخرطوم. مسترجع من http://search.mandumah.com/Record/916054

إسلوب MLA
عبدالله، سمية عبدالله محمد الأمين، و شاذلى صديق محمد أحمد. "التنبؤ بالمبيعات بإستخدام شجرة القرار: بالتطبيق على شركة سويتش للتجارة والهندسة قسم المولدات" رسالة ماجستير. جامعة النيلين، الخرطوم، 2017. مسترجع من http://search.mandumah.com/Record/916054

© 2024 المنظومة. جميع الحقوق محفوظة.
هذه المادة متاحة بناء على الإتفاق الموقع مع أصحاب حقوق النشر، علما أن جميع حقوق النشر محفوظة. يمكنك تحميل أو طباعة هذه المادة للاستخدام الشخصي فقط، ويمنع النسخ أو التحويل أو النشر عبر أي وسيلة (مثل مواقع الانترنت أو البريد الالكتروني) دون تصريح خطي من أصحاب حقوق النشر أو المنظومة.

الفصل الخامس

(النتائج والتوصيات)

5-1 مناقشة النتائج :

```
=== Run information ===

Scheme:       weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:     switchdata
Instances:    1000
Attributes:   9
              NO
              model of genarator
              voltage
              Qty
              customer
               Email
              mobile
              description of genarator
              date
Test mode:    split 66.0% train, remainder test

=== Classifier model (full training set) ===
```

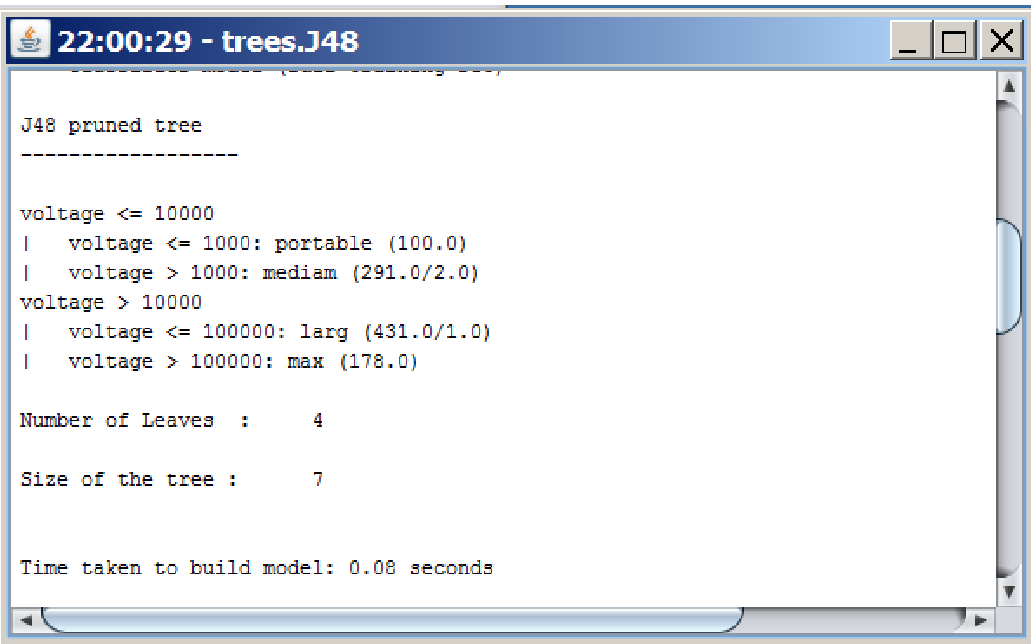
شكل رقم(29) يوضح معلومات التنفيذ

هذا الشكل يحتوي على : اسم الخوارزمية المستخدمة وهي J48 , اسم العلاقة (الجدول) وهو switchdata , عدد السجلات = 1000 , عدد الخصائص = 9 وهي

(no)رقم العملية الشرائية) –modle of genarator) موديل المولد) –voltage)سعة المولد الفولتية)–Qty (كمية المولدات المشتراه بواسطة زبون معين)–customer(اسم الجهة او الزبون الذي قام بعمليه الشراء) –mobile(رقم محمول لجهة الشراء)_Email (بريد الكتروني لجهة الشراء)

– date (تاريخ عملية الشراء)–description of generator (وصف المولد من حيث الحجم)

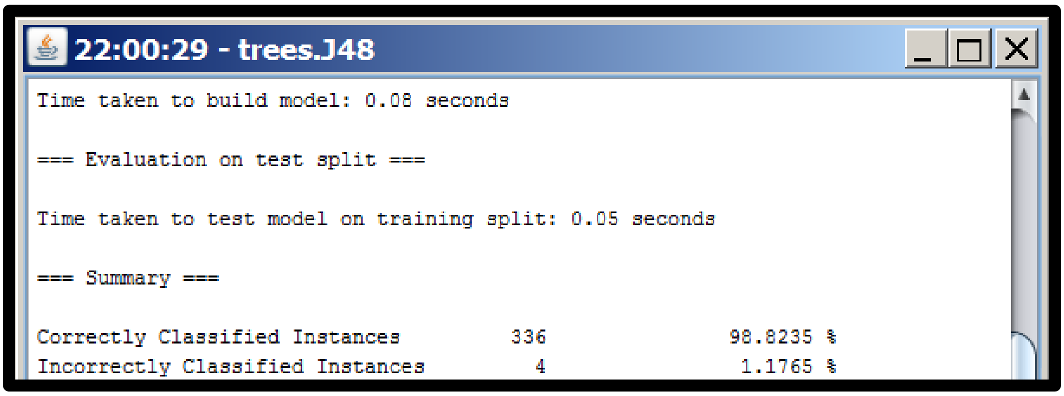
نمط الاختبار المختار هو Split66% يعني أن 66% من مجموعة البيانات تم أخذها لعملية التدريب و34 % من مجموعة البيانات المتبقية تم أخذها لعملية الاختبار.



```
22:00:29 - trees.J48

J48 pruned tree
------------------

voltage <= 10000
|   voltage <= 1000: portable (100.0)
|   voltage > 1000: mediam (291.0/2.0)
voltage > 10000
|   voltage <= 100000: larg (431.0/1.0)
|   voltage > 100000: max (178.0)

Number of Leaves  :     4

Size of the tree :      7


Time taken to build model: 0.08 seconds
```

شكل رقم (30) يوضح نموذج المصنف

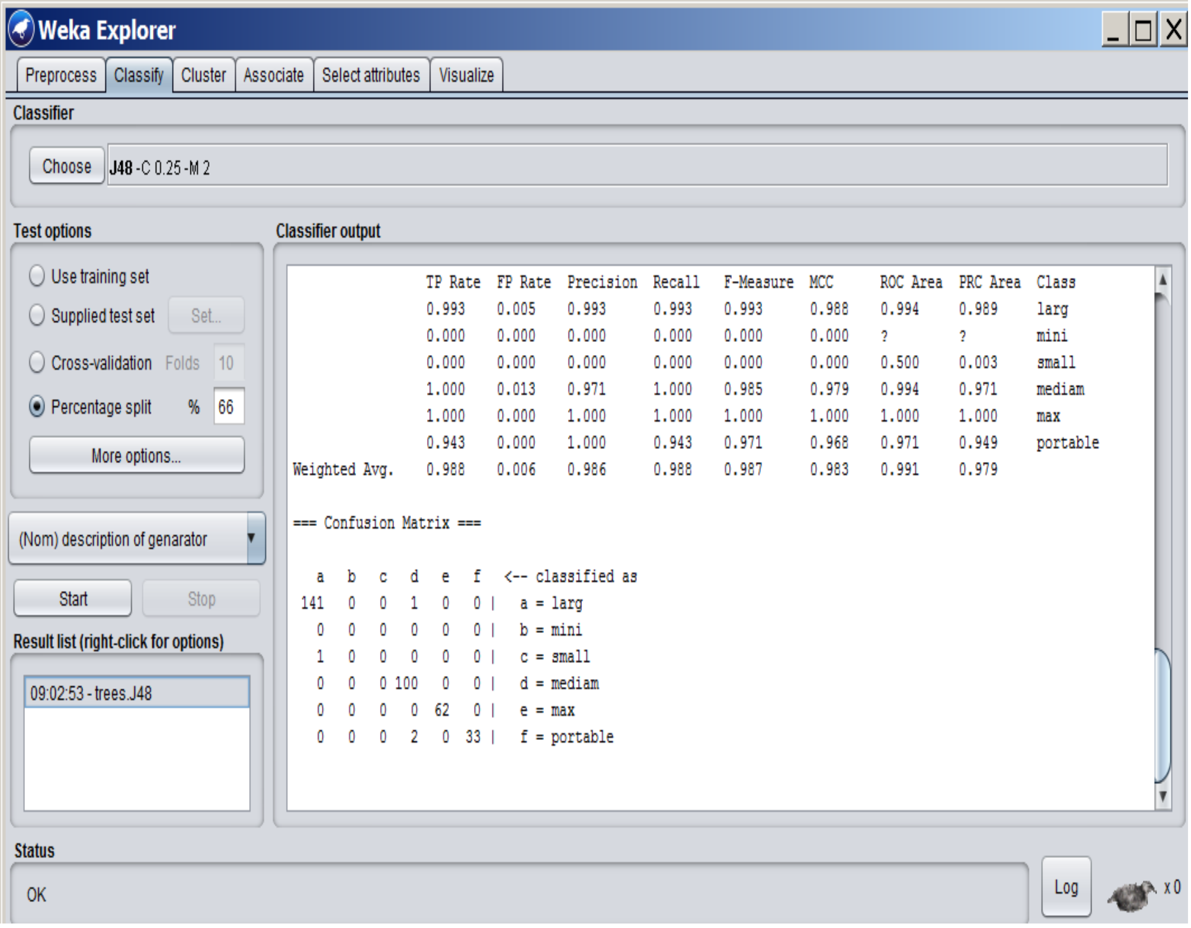
هذا الشكل يوضح نموذج المصنف الذي تم بناءه وهو شجرة قرار فيها عدد الأوراق = 4, وعدد العقد في الشجرة (حجم الشجرة) = 7, الزمن الذي أخذ في بناء النموذج هو 0.08 من الثواني.



```
22:00:29 - trees.J48
Time taken to build model: 0.08 seconds

=== Evaluation on test split ===

Time taken to test model on training split: 0.05 seconds

=== Summary ===

Correctly Classified Instances         336               98.8235 %
Incorrectly Classified Instances         4                1.1765 %
```

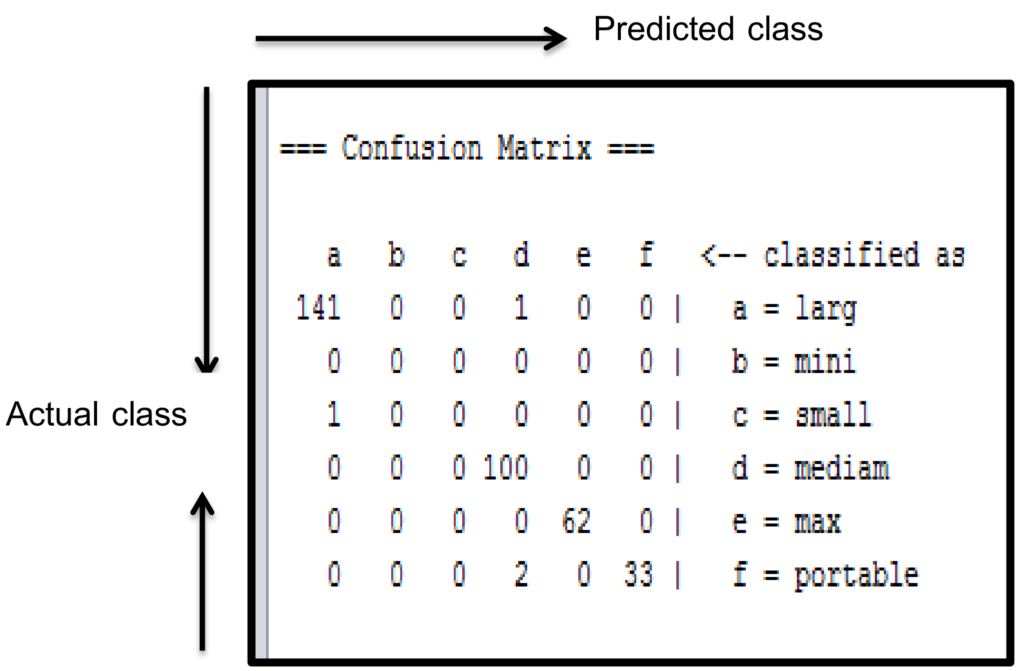
شكل رقم (شكل31) يوضح تقييم المصنف على بيانات الاختبار

هذا الشكل يعطينا تخمينات أداء الشجرة التنبؤي الذي أنتج بوحدة تقييم WEKA , ما هي دقة المصنف الذي كان قادر على توقع الصنف الحقيقي للحالات تحت وحدة الاختبار المختارة حيث كانت نسبة الحالات التي صنفت بشكل صحيح تقريبا 98.8% والحالات التي صنفت بشكل خاطئ تقريبا 1.18 %.



شكل رقم (32) يوضح مصفوفة التضارب

هذه المصفوفة تسمى ايضا مصفوفة الخطأ أو :(Confusion Matrix) مصفوفة التضارب وهي مصفوفة مشهورة في مجال تعليم الآلة, تستخدم لاختبار أداء الخوارزميات. وهي عبارة عن جدول يحوي على معلومات وتفاصيل حول التصنيفات الفعلية (التي تم تصنيفها من قبل الإنسان) والتصنيفات التنبئية التي تنبؤها المصنف. فكل عامود في المصفوفة يمثل الفئات المُتنبئة (Predicted class)و كل صف يمثل الفئات الفعلية. (Actual class) ويتم تقييم أداء المصنف عادةً باستخدام البيانات الواردة في المصفوفة, ويكون حجم المصفوفة حسب عدد الفئات.




```
=== Confusion Matrix ===

   a   b   c   d   e   f   <-- classified as
 141   0   0   1   0   0 |   a = larg
   0   0   0   0   0   0 |   b = mini
   1   0   0   0   0   0 |   c = small
   0   0   0 100   0   0 |   d = mediam
   0   0   0   0  62   0 |   e = max
   0   0   0   2   0  33 |   f = portable
```

شكل رقم (33) مصفوفة التضارب الناتجة لعملية اختبار المصنف

الشكل اعلاه يوضح مصفوفة التضارب الناتجة لعملية اختبار المصنف على بيانات الاختبار 34 % , مصفوفة التضارب هي أداة مفيدة لتحليل كيف يتعرف المصنف بصورة جيدة على السجلات للأصناف المختلفة.

من مصفوفة التضارب نحسب الأتي :

Accuracy = (141 + 100+62+33) / (141 + 1+1+100+62+2+33)

= (336) / (340) = 0.9882352941176471 = 98.82 %

Error rate = (1 + 1+2) / (141 + 1+1+100+62+2+33)

= (4) / (340) = 0.0117647058823529= 1, 18 %

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.993	0.005	0.993	0.993	0.993	0.988	0.994	0.989	larg
	0.000	0.000	0.000	0.000	0.000	0.000	?	?	mini
	0.000	0.000	0.000	0.000	0.000	0.000	0.500	0.003	small
	1.000	0.013	0.971	1.000	0.985	0.979	0.994	0.971	mediam
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	max
	0.943	0.000	1.000	0.943	0.971	0.968	0.971	0.949	portable
Weighted Avg.	0.988	0.006	0.986	0.988	0.987	0.983	0.991	0.979	

شكل رقم (34) النتائج النهائية

Accuracy=98.8%

Error rate=1.18%

حيث ان

1. TP(True Positive): عدد الحالات التي تنبئها المصنف بشكل صحيح لفئة p و تنتمي بشكل فعلي لفئةp
2. FN(False Negative): عدد الحالات التي تنبئها المصنف بشكل غير صحيح لفئة p و تنتمي بشكل فعلي لفئةn
3. TN(True Negative): عدد الحالات التي تنبئها المصنف بشكل صحيح لفئة n و تنتمي بشكل فعلي لفئةp
4. FP(False Positive): عدد الحالات التي تنبئها المصنف بشكل غير صحيح لفئة n و تنتمي بشكل فعلي لفئةp
5. نسبة الضبطية Precision : نسبة السجلات التي صنفها المصنف على أنها Positive وهي في الحقيقة Positive.
6. نسبة الاستدعاء Recall : نسبة السجلات ال Positive والتي صنفها المصنف على أنها Positive.

larg القيمة

True Positive Rate (**TPR**) = TP/(TP + FN)

= 141 /(141+1) = 0.9929577464788732 =0.993

mini القيمة

True Positive Rate (**TPR**) = TP/(TP + FN)

= 0 /(0+0) = 0

small القيمة

True Positive Rate (**TPR**) = TP/(TP + FN)

= 0 /(1+0) = 0

mediam القيمة

True Positive Rate (**TPR**) = TP/(TP + FN)

= 100 /(100+0) = 1 =1

max القيمة

True Positive Rate (**TPR**) = TP/(TP + FN)

= 62 /(62+0) = 1

portable القيمة

True Positive Rate (**TPR**) = TP/(TP + FN)

= 33 /(33+2) = 0.9428571428571429=0.943

5-2 النتائج النهائية :

من خلال دراستنا لمبيعات شركة سويتش للاعوام من 2015–2017 باستخدام تقنيات التنقيب، توصلنا للنتائج التالية:

1. تم التوصل الى درجة دقة بنسبة % 98.82 ونسبة خطأ % 1.18 وهذا يدل علي تحليل البيانات بالصورة المطلوبة اذ تعتبر نسبة الدقة عاليه
2. كما تم التوصل للتالي :

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.993	0.005	0.993	0.993	0.993	0.988	0.994	0.989	larg
	0.000	0.000	0.000	0.000	0.000	0.000	?	?	mini
	0.000	0.000	0.000	0.000	0.000	0.000	0.500	0.003	small
	1.000	0.013	0.971	1.000	0.985	0.979	0.994	0.971	mediam
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	max
	0.943	0.000	1.000	0.943	0.971	0.968	0.971	0.949	portable
Weighted Avg.	0.988	0.006	0.986	0.988	0.987	0.983	0.991	0.979	

5-3التوصيات:

من خلال استخدامنا لتقنيات تنقيب البيانات في التسويق الذكي،نوصي بالتالي:

1. تنشيط إدارة التسويق بالشركة واحاطتها بهذا العلم للاستفادة منه في تحليل بياناتها وتحليل السوق.
2. لإعطاء نتائج تفيد في الحملات التسويقية بصورة افضل يجب على إدارة المبيعات ادخال البيانات بصورة صحيحة و استخدام أكبر قدر من البيانات.
3. البريد الالكتروني عنصر اساسي داعم للتسويق لانه القاسم المشترك لمعظم وسائل التواصل الاجتماعي وكذلك التلفون يجب مراعاه ادخالها
4. عمل استبيان للزبائن الذين تم شراؤهم لسعات معينة من المولدات بتقدير شخصي وليس استنادا على تقدير مهندسي الشركة لمعرفة اسباب شرائهم لتلك السعات وتنويرهم بالانسب .
5. معرفة ومواكبة الجديد من برامج التنقيب وجدوى استخدامها لدى الشركة

5-4الخاتمة :

الحمد لله حمدا كثيرا طيبا مباركا فيه على اكمال هذا البحث فان تم العمل به والاستفادة من توصياته في الشركة وتم العمل بموجب هذا التحليل في المستقبل فهذا ما ارجوه وان لم يتم العمل به فهو علم عساه ان يرى النور يوما لشركات اخرى.

في هذه الدراسة تم استخدام واحدة من أهم تقنيات تنقيب البيانات وهي خوارزمية شجرة القرار (J48) , حيث تم تطبيقها على مجموعة بيانات مبيعات شركة سويتش للتجارة والهندسة قسم المولدات 2015-2017 حيث كانت نسبة الدقة هي 98.82 % وتوصلنا إلى أنه يجب ان يتم التركيز على السعات larg في الاستيراد وذلك لزيادة الطلب عليها , ويجب ان يتم التركيز والتنوير الاعلاني في المستقبل على المولدات ذات السعات portable .

ولله الحمد والشكر والصلاة والسلام على خير الانام سيدنا مُحَمَّد ﷺ وعلى آله وصحبه اجمعين,